We claim:

1. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of:

receiving at least one protein backbone structure;

applying a protein design algorithm to generate a protein sequence and structure;

sampling and evaluating one or more amino acids and rotamers within the context of said protein sequence and structure;

generating a probability matrix for said amino acids and rotamers that represent the viable sequence space for said protein backbone.

2. A method according to claim 1 further comprising the step of:

generating a single protein sequence from said probability matrix.

3. A method according to claim 1 further comprising the step of:

generating a combinatorial library of proteins from said probability matrix.

4. A method according to claim 1 wherein said steps are repeated more than once to generate said probability matrix

5. A method according to claim 1 wherein said protein design algorithm comprises an optimization procedure selected from the group of: dead end elimination algorithm; genetic algorithm; Monte Carlo algorithm; and self consistent mean field theory algorithm or combinations thereof.

6. A method according to claim 1 wherein said protein backbone structure is taken from a natural protein.

7. A method according to claim 1 wherein said protein structure is generated by comparative modeling.

8. A method according to claim 1 wherein the information from at least two probability matrices is combined to satisfy at least two constraints on sequence space.

9. A method according to claim 1 wherein said protein backbone structure comprises an ensemble of related protein backbone structures.

10. A method according to claim 9 further comprising the step of:

generating a single protein sequence from said probability matrix.

11. A method according to claim 9 further comprising the step of:

generating a combinatorial library of proteins from said probability matrix

12. A method according to claim 9 wherein said steps are repeated more than once to generate said probability matrix.

13. A method according to claim 9 wherein said protein design algorithm comprises an optimization procedure selected from the group of: dead end elimination algorithm; genetic algorithm; Monte Carlo algorithm; and self consistent mean field theory algorithm or combinations thereof.

14 A method according to claim 9 wherein said ensemble of related protein backbone structures are taken from a family of natural proteins.

15. A method according to claim 9 wherein said ensemble of related backbone structures is derived from an NMR structure.

16. A method according to claim 9 wherein said ensemble of related protein backbone structures is generated by a Monte Carlo simulation.

17. A method according to claim 9 wherein said ensemble of related protein backbone structures is generated by a molecular dynamics simulation.

18. A method according to claim 9 wherein the information from at least two probability matrices is combined to satisfy at least two constraints on sequence space.

19. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of:

receiving at least one complete protein sequence and structure;

sampling and evaluating one or more amino acids and rotamers

within the context of said protein sequence and structure;

generating a probability matrix for said amino acids and rotamers that represent the viable sequence space for said protein backbone.

20. A method according to claim 19 wherein said protein sequence an structure is that of a natural protein.

21. A method according to claim 19 wherein said protein sequence and structure comprises an ensemble of related protein structures.

22. A method according to claim 21 wherein said ensemble of proteins is generated by a Monte Carlo simulation.

23. A method according to claim 21 wherein said ensemble of proteins is generated by a molecular dynamics simulation .

24. A method according to claim 19 wherein said steps are repeated more than once to generate said probability matrix.

25. A method according to claim 19 further comprising the step of :

generating a single protein sequence from said probability matrix.

26. A method according to claim 19 further comprising the step of:

generating a combinatorial library of proteins from said probability matrix.

27. A method according to claim 19 wherein said protein sequence and structure is generated by comparative modeling.

28. A method according to claim 19 wherein said protein sequence and structure is taken from a natural protein.

29. A method according to claim 19 wherein the information from at least two probability matrices is combined to satisfy at least two constraints on sequence space.

30. A method for optimizing simulation or scoring function parameters that utilizes comparisons between designed sequences and natural sequences, comprising the steps of:

designing a protein sequence;

comparing said designed protein sequence to natural protein statistics;

modifying said simulation or scoring function parameters consistent with said comparison.

31. A method according to claim 30 wherein said steps are repeated at least once.

32. A method according to claim 30 wherein said natural protein statistics are in the form of a position specific scoring matrix.

33. A method according to claim 30 wherein said natural protein

statistics are in the form of amino acid composition

34. A method for optimizing simulation or scoring function parameters that utilizes comparisons between designed sequences and natural sequences, comprising the steps of:

calculating an amino acid probability matrix;

comparing said matrix to natural protein statistics;

modifying simulation or scoring function parameters consistent with said comparison.

35. A method according to claim 34 wherein the sequence of steps is repeated at least once.

36. A method according to claim 34 wherein said natural sequence statistics are in the form of a position specific scoring matrix.

37. A method according to claim 34 wherein said natural sequence statistics are in the form of amino acid composition.